THE SUITABILITY OF CONSUMERS' ASSESSMENTS OF PHYSICIAN AND HOSPITAL PERFORMANCE AS INDICATORS OF THE QUALITY OF CARE

John E. Ware, Jr., Ph.D., Allyson Ross Davies, Ph.D., and Haya R. Rubin, M.D.

The RAND Corporation and UCLA School of Medicine

December 1987

Contractor Document
Health Program, Office of Technology Assessment
U.S. Congress, Washington, DC 20510

This paper was prepared by an outside contractor for the OTA assessment The Quality of Medical Care: Information for Consumers. The paper does not necessarily reflect the analytical findings of OTA, the assessment's advisory panel, or the Technology Assessment Board.

J952-76

ACKNOWLEDGMENTS

We would like to acknowledge Leila Baladi and Marilyn Weinstein, who provided valuable assistance in identifying and maintaining a database for the literature relevant to this review.

Our thanks also go to reviewers of an earlier draft, who provided useful substantive and editorial comments. They include: Paul B. Batalden, Robert H. Brook, Paul D. Cleary, Denise Dougherty, Peter Goldschmidt, Judith A. Hall, Floyd Loop, Harold Luft, Fred Mosteller, and Eugene C. Nelson. A special of thanks goes to Jane Sisk, our project officer at OTA.

Chapter 1

INTRODUCTION

History indicates that efforts to assess the quality of medical care have waxed and waned (Brook and Avery, 1976). Current interest in quality of care assessment and assurance (QA) activities is fueled in large part by concerns that the cost-containment policies of the early 1980s were implemented with little attention to their effects on the quality of services. Along with this resurgence of interest in QA comes renewed attention to the consumer's viewpoint, and efforts to involve consumers in QA activities are likely to increase rapidly. The Office of Technology Assessment's (OTA's) interest in the suitability of data provided by consumers as indicators of quality of care for individual physicians and hospitals is a case in point. Although several factors may motivate decisions to involve the consumer in such activities, one assumption appears to be critical: that consumers can provide valid information about the quality of their medical care.

"Quality of care" refers to the worth or excellence of various attributes of medical care. Donabedian (1980) has proposed that quality be judged relative to "what is known about the relationship between the characteristics of the medical care process and their consequences to ...health." He further identifies two basic characteristics of the medical care process: technical and interpersonal. He notes that while professionals generally define the norms against which the quality of technical process is judged, the "values...and rules that govern relationships among people" provide the standards for judging quality of the interpersonal aspects of medical care. Most definitions of quality that focus on a single encounter or an episode of treatment include both attributes (Donabedian, 1966, 1980; Brook

and Avery, 1977; American Board of Internal Medicine, 1983). Broader definitions, pertinent to evaluating delivery systems, might include such attributes as the accessibility and availability of providers and services. These attributes certainly reflect the quality of system performance, and they can be important determinants of outcomes of care (Brook et al., 1983; Ware et al., 1986; Lurie et al., 1984, 1986). Recognizing OTA's interest in the quality of institutional or system performance as well as that of individual providers, we use the broader definition in this report.

There are at least three good reasons to seek input from consumers when evaluating quality of medical care. First, it ensures that those evaluations will represent the values of the consumers of services. Second, consumers are the only source of information regarding certain features of medical care (i.e., interpersonal aspects of the provider-patient relationship), and may provide valid information that fills gaps in other, more traditional sources of information on quality, such as medical records (Lebow, 1974; Lyons and Payne, 1974; Christoffel and Loewenthal, 1977; Thompson and Osborne, 1976; Gerbert and Hargreaves, 1987; Ware et al., forthcoming). Thus, consumers can be both unique and supplementary sources of data on quality-related attributes. Third, consumer surveys cost no more and probably less than data obtained from more traditional sources (Davies and Ware, in press; see also Chapter 4), and thus going to consumers for data on those quality indicators for which they provide valid information should help achieve a cost-effective evaluation.

We have reviewed the published literature to identify those quality-related attributes of care (i.e., accessibility and availability; technical process; interpersonal aspects), both inpatient and outpatient, on which consumers can provide valid information. If valid, we would expect consumers'

assessments to correspond with known differences in these attributes. Where empirical evidence of validity is sparse or lacking, we identify the types of information needed about data from consumers to warrant its use in QA activities. We also address practical considerations involved in obtaining data from consumers for purposes of evaluating the quality of physician and hospital care.

Chapter 2

GENERIC ISSUES

We identify and discuss briefly in this chapter several distinctions that were important to our review of the suitability of data from consumers as indicators of quality of care.

OUTPATIENT VS. INPATIENT SETTING

The potential of and problems in using information from consumer assessments as indicators of quality differ in a number of important respects between outpatient and inpatient care settings. Research relevant to quality of care that uses data from consumers has proceeded largely independently and at different paces in the two settings, and both the state-of-the-art of concept definition and measurement development and validation is much further advanced for consumer-based quality indicators in outpatient than inpatient settings. The attributes of in outpatient care distinguished by consumers correspond well with those discussed in the Introduction (i.e., access, interpersonal aspects, technical process); attributes of inpatient care appear to more closely replicate the institutional framework. Finally, whether outpatient or inpatient care settings are the subject of study is important to the conclusions that are drawn about the validity of consumer assessment and to judgments regarding the advantages and disadvantages of consumer assessments relative to other data sources (e.g., medical record abstractions). For these reasons, we generally discuss evidence relevant to consumer data on outpatient and inpatient care separately.

Attributes of Quality

Conclusions about the validity of consumer information are likely to vary according to the attribute of quality being evaluated. Consumers are more qualified to judge the interpersonal than the technical quality of care. Alternative sources of data about quality clearly favor the technical process of care. The amount of evidence regarding the validity of data from consumers as quality indicators varies greatly by attribute; most studies have focused on interpersonal aspects in outpatient settings. Because different attributes have been hypothesized in studies of inpatient and outpatient settings, the validity variables involved vary accordingly. Finally, unlike the case in literature on outpatient settings, the literature on inpatient care lacks a coherent taxonomy of quality from the consumer's perspective. Consequently, while research in outpatient settings has focused on the specific attributes of quality identified in Chapter 1, many consumer-based indicators used in inpatient settings represent considerable aggregation across attributes (e.g., an assessment of quality in general). Thus, while we have organized our review according to three attributes of quality--access, interpersonal aspects, technical process--we note that this framework better fits our summary of literature on consumer indicators of outpatient quality. Recognizing the nature of validation studies on inpatient indicators, we have included a fourth attribute--overall quality--in our review of that evidence.

Ratings vs. Reports

Consumers' ratings of their care must be distinguished from their reports about care. Consumers' ratings capture personal evaluations of attributes of medical care providers and services; they are inherently more subjective because they reflect personal experiences, expectations and

preferences, as well as the standards consumers apply when evaluating care.[1] By contrast, consumers' reports deal with things that did or did not occur; they are inherently more objective and can be more readily confirmed by an outside observer.

SUMMARY

Based on the preceding distinctions, we note that the validity of consumer data is likely to vary depending on whether ratings and reports are obtained, the attribute of quality considered, and the setting involved. Table 2.1 illustrates these distinctions with examples of ratings and reports of three quality-related aspects of an episode of medical care, inpatient or outpatient: access, interpersonal aspects, and technical process. These distinctions, which guided our review, should be kept in mind when reading the remainder of this report.

---

1 Most of the studies we reviewed used attitudinal measures, or more specifically, patient satisfaction measures, to obtain data from consumers. The majority of items in these surveys can be considered evaluations, either because the respondent is asked the strength of their endorsement of an evaluative statement (e.g., "My doctors are very competent and well-trained"), or because the response categories offered constitute an evaluation (e.g., "excellent" to "poor"; "satisfied" to "dissatisfied).

Table 2.1--Distinction Between Consumers' Ratings and Reports Regarding Quality Related Attributes of Medical Care

Attribute	Rating	Report
Technical	Evaluation (excellent-poor of doctor's thoroughness	Indication (yes-no) of whether doctor checked blood pressure
Interpersonal	Evaluation (excellent-poor) of courtesy and respect shown by doctor	Indication (yes-no) of whether doctor introduced self
Accessibility	Evaluation (excellent-poor) of office waiting time	Indication (yes-no) of whether office wait was 15 minutes or less

Chapter 3

METHODS

Identification of Publications

This review covered articles, reports, and books published through summer 1987. We focused on the measurement and meaning of consumers' assessments of the quality of medical, surgical, psychiatric, nursing, and obstetric care in both outpatient and inpatient settings. To identify relevant articles, we used a combination of four strategies: (1) articles identified during an earlier review that covered both literature on both outpatient and inpatient care through 1976 (Ware et al., 1978) and a recent review of literature on satisfaction with inpatient care (Rubin et al., forthcoming); (2) computerized searches based on relevant key words (3) backward searches of citations in articles identified at steps 1 and 2; and (4) review of titles listed in tables of content for selected journals published over the past 10 years.

Selection of Publications for In-depth Review

This approach identified approximately 450 publications on the subject of patient assessments of and/or reports about their medical care, both outpatient and inpatient. Given the number of publications, we screened them to identify those that were most likely to provide relevant information and thus that we would abstract in detail. In screening studies for in-depth review, we placed greatest emphasis on studies that, from a methodological perspective, had strong designs and provided adequate information about interventions (if any), reliability of consumer data, and operational

definitions of the variables studied in relation to consumer information. In particular, we favored studies that formally tested for a direct link between actual differences in quality and consumer ratings or reports, either by manipulating quality experimentally or by obtaining measures of actual quality independent of those provided by consumers. In the case of experiments, we required some form of manipulation check to determine that quality had actually been altered.

We were most interested in whether consumers' ratings and reports of a given feature of care reflected known differences in that feature and not others. Measures used to quantify such differences are referred to as validity variables. We also favored publications that reported analytic methods and findings in sufficient detail to know whether the association between a consumer assessment of quality and a validity variable was statistically significant. (See below for more detailed definitions of reliability and validity.)

In reviewing the considerable literature on outpatient care, we excluded studies that used as validity variables interventions or measures that did not distinguish specific attributes of quality. Because the development and validation of consumers' ratings of inpatient care has lagged behind that of outpatient care, we were willing to consider evidence from studies that linked measures of "overall" quality (e.g., a mix of interpersonal and technical features) to consumer assessments.

Following these criteria, we selected 33 articles pertinent to the validity of information from consumers about quality of care care for in-depth review and have summarized findings in Chapter 4. Of the 33, 26 publications contained information relevant to the validity of consumer information about

quality-related aspects of outpatient care, and 7, inpatient care. Table 3.1 indicates the number of publications included in the in-depth review of validity evidence by setting, attribute of care, and type of information. (Entries in Table 3.1 do not sum to 33 because several studies reported results pertinent to validity of consumer information as indicators of more than one quality-related attribute.)

Table 3.1--Studies of Validity of Consumer Information Regarding Quality of Medical Care, by Setting, Attribute of Care, and Type of Information

Setting	Number of Studies, by Attribute and Type of Information	
Outpatient	Access	
	Ratings	5
	Reports	0
	Interpersonal Aspects	
	Ratings	17
	Reports	0
	Technical Process	
	Ratings	9
	Reports	3
Inpatient	Access	
	Ratings	0
	Reports	0
	Interpersonal Aspects	
	Ratings	3
	Reports	0
	Technical Process	
	Ratings	2
	Reports	3

Nine of these studies tested for the effects of experimental manipulation of quality-related features of care on consumer assessments. Another nine used observational methods to collect data on validity variables (e.g., videotapes of provider-patient encounters) and described elements of the encounters according to objective coding schemes. Seven studies relied on provider report and/or chart review to obtain data for validity variables. The remaining studies (as well as some of the preceding ones) used other sources of information (e.g., data on physician/population ratios, staff ratings of ward performance) as validity variables. Of the 33 studies that we reviewed in depth for validity information, 7 also provided reliability estimates for consumers' ratings. We identified another 20 studies that contained estimates of the reliability of consumers' ratings of quality-related features of inpatient and outpatient medical care.

Definitions of Reliability and Validity

Reliability is the the proportion of measured variance that is true score, as opposed to random error. A reliability estimate can be obtained by correlating scores on two forms of a measure (alternate-forms), scores for the same measure at two points in time (test-retest), or scores on items measuring the same concept (internal-consistency). In each case, the estimate ranges from 0.0 to 1.0. Generally, minimum acceptable standards for reliability range from 0.50 (Helmstadter, 1964) to 0.70 (Nunnally, 1978) for measures to be used in group comparisons (e.g., patient samples from two outpatient clinics or two hospitals). Most uses of consumer information on quality-related topics will involve group, rather than individual, comparisons.

A valid measure contains information about the variable of interest. By focusing on validity, one distinguishes between any information and the desired information. Many approaches can be used to evaluate the validity of a rating or report as an indicator of quality. Their goals are to: (1) determine the relationship between scores for the quality indicator in question and the chosen validity variable (convergent evidence of validity); and (2) demonstrate that the indicator in question has a weaker relationship with measures of other attributes (discriminant evidence of validity). Because there is no "gold standard," or no one indicator of the "true" quality of medical care, studies pertinent to the validity of consumer information about quality of care rely on proxy indicators that vary depending on the quality-related attribute of care being considered. To illustrate, one kind of validity variable would be appropriate for validating consumers' assessments of access (e.g., measures of actual office waiting times); another would be appropriate for validating a consumer assessment of technical quality (e.g., independent observation of the process of diagnosis and management). Validity standards were applied in a manner consistent with generally accepted guidelines (Standards..., 1974).

CHAPTER 4

FINDINGS

If data from consumers are to be used as indicators of quality, the data must be reliable, valid, and feasible to obtain. After summarizing evidence related to reliability of information from consumers, we turn to the focal point of this review: evidence regarding the validity of that information. In reviewing evidence of validity, we consider consumer data on outpatient and inpatient care separately. We conclude with comments regarding the feasibility of obtaining data from consumers for use as quality indicators.

Reliability

Despite their importance, the majority of studies we identified did not report reliability estimates. Among studies of consumers' ratings in outpatient settings, 21 included reliability estimates, as did 8 studies of ratings in inpatient settings; several reported estimates for more than one sample. When reliability estimates were reported, they were above 0.50 for virtually all multi-item rating scales, and many exceeded the 0.70 standard (see Table 4.1).[1] This finding holds for many relatively short (e.g., fewer than 10 items per concept) but well-constructed multi-item measures, even in disadvantaged populations where reliability tends to be poorer (e.g., Hulka et

1 Several of the studies in Table 4.1 did not meet our criteria for indepth review for information relevant to validity; they are included here because they do provide data regarding the reliability of consumer data on quality-relevant features of care.

Table 4.1--Summary of Information About Reliability of Consumer Ratings of Outpatient and Inpatient Medical Care

Author/Year (Sample)	Estimation Method [a]	Results by Scale: Estimate (No. of Items) [b]	
Ratings of Outpatient Medical Care			
Franklin & McLemore 1967, 1970 (n=136)	IC	General Satisfaction	.87 (20)
Hulka et al. 1970 (n=49)	AF	Professional Competence	.63 (12-14)
		Personal Qualities	.75 (12-14)
		Access/Finances	.43 (12-14)
		Total	.80 (42)
Zyzanski et al. 1974 (n=426)	IC	Professional Competence	.75 (14)
		Personal Qualities	.86 (14)
		Access/Finances	.68 (14)
		Total	.90 (42)
Rojek et al. 1975 (n=1100)	IC	General Satisfaction	.71 (3)
Aday & Andersen 1975 (n=>2000)	IC	Access/Finances	.68 (3)
		Interpersonal/Technical Quality	.90 (8)
		Total	.84 (11)
Ware et al. 1975 (n=903)	IC	Availability Total	.83 (10)
		Continuity	.78 (4)
		Finances	.69 (4)
		Interpersonal	.67 (3)
		Technical Quality	.89 (4)
		Interpersonal/Technical Total	.89 (25)
Ware and Snyder 1975 (n=433)	IC [c]	Availability (4)	.47-.76 (2)
		Accessibility (3)	.49-.64 (2-3)
		Continuity (2)	.57-.67 (2)
		Finances (3)	.66-.75 (2)
		Interpersonal (3)	.67-.75 (3-4)
		Technical Quality (5)	.52-.73 (2-4)

Table 4.1--Summary of Information About Reliability of Consumer Ratings of Outpatient and Inpatient Medical Care (cont'd)

Author/Year (Sample)	Estimation Method [a]	Results by Scale: Estimate (No. of Items) [b]	
Ratings of Outpatient Medical Care (cont'd.)			
Ware and Snyder 1975 (n=167)	TRT [c] (6-week interval)	Availability (4)	.57-.62 (2)
		Accessiblity (3)	.59-.62 (2-3)
		Continuity (2)	.59-.64 (2)
		Finances (3)	.62-.69 (2)
		Interpersonal (3)	.62-.69 (3-4)
		Technical Quality (5)	.64-.70 (2-4)
Roter 1977 (n=250)	IC	Overall Satisfaction	.67 (6)
DiMatteo et al. 1980 (n=4 to 10 pts per MD, 29 MDs)	IC (across pts. for ind. MD)	Interpersonal Aspects 10 pts/MD)	.61 (NA)
		Interpersonal Aspects (4-5 pts/MD)	.12 (NA)
Breslau et al. 1981 (n=370)	IC	Access	[d] (8)
		Availability of Resources	[d] (5)
		Continuity	[d] (3)
		Finances	[d] (6)
		Humaneness	[d] (8)
		Technical Quality	[d] (7)
Ware et al. 1981 (n=2287)	IC	Technical Quality	.70 (4)
		Interpersonal Aspects	.66 (3)
		General Satisfaction	.74 (4)
		Access to Care	.74 (7)
Marquis et al. 1983 (n=279)	IC	General Satisfaction	.70 (4)
Bartlett et al. 1984 (n=60)	IC	Overall Satisfaction	.88 (8)
Chang et al. 1984 (n=268)	IC	Global Satisfaction	.95 (7)
Corah et al. 1984 (n=24)	IC	Information/Communication	.94 (3)
		Understanding/Acceptance	.87 (3)
		Technical Competence	.84 (4)
		Total	.84 (10)
DiMatteo et al. 1986 (n=329)	IC	Communication	.72 (8)
		Affective Care	.79 (9)
		Technical Care	.65 (3)

Table 4.1--Summary of Information About Reliability of Consumer Ratings of Outpatient and Inpatient Medical Care (cont'd)

Author/Year (Sample)	Estimation Method [a]	Results by Scale: Estimate (No. of Items) [b]	
Ratings of Outpatient Medical Care (cont'd.)			
DiMatteo et al. 1986 (n=6-7 pts per MD, 57 MDs)	IC (across pts. for ind. MD)	Communication	.52 (8)
		Affective Care	.46 (9)
		Technical Care	.40 (3)
Cope et al. 1986 (n=424)	IC	Art of Care	.92 (9)
		Technical Quality	.81 (5)
Davies et al. forthcoming (n=1537)	IC [e]	Access Total	.65-.70 (8)
		Avail./Family MDs	.73-.78 (2)
		Avail./Hospitals	.78-.84 (3)
		Availability Total	.64-.74 (5)
		Costs of Care	.71-.80 (2)
		Interpersonal Aspects	.80-.83 (8)
		Technical Quality	.65-.72 (6)
		Facilities	.78-.83 (2)
		Quality Total	.89-.91 (16)
		General Satisfaction	.66-.75 (4)
Ware et al. forthcoming (n=109)	IC	Interpersonal Aspects	.93 (5)
		Technical Quality	.90 (5)
Ratings of Inpatient Medical Care			
Rice et al. 1963 (n=457)	TRT (1 week)	Physical Facilities	.78 (22)
		Patient Services	.77 (27)
		Patient Management	.67 (20)
		Ward Evaluation Scale	.81 (69)
Souelem 1955 (n=95)	AF	General Attitudes Toward Mental Hospitals	.88 (36)
Hinshaw & Atwood 1982 (5 studies, n's ranged from 49 to 237)	IC	Technical/Prof.	.64-.97 (NA)
		Trusting Relationship	.82-.92 (NA)
		Education	.49-.95 (NA)
		Total	.89 (NA)
Wales et al. 1983 (n=115)	IC	General	.82 (NA)
		Competency	.73 (NA)
		Humaneness	.72 (NA)
		Total	.88 (NA)
		Physical Environment	.80 (NA)

Table 4.1--Summary of Information About Reliability of Consumer Ratings of Outpatient and Inpatient Medical Care (cont'd)

Author/Year (Sample)	Estimation Method[a]	Results by Scale: Estimate (No. of Items)[b]	
Wales et al. 1983 (n=115)	TRT [f] (24-hour)	General	.87,.92 (NA)
		Competency	.85,.84 (NA)
		Humaneness	.83,.83 (NA)
		Total	.93,.91 (NA)
		Physical Environment	.92,.86 (NA)
Carmel 1985 (n=476)	IC	Physicians	.94 (11)
		Nurses	.95 (11)
		Supportive Services	.86 (9)
Greenley 1985 (n=177)	IC	Humaneness of Staff	.85 (NA)
		Humaneness of Psychiatrist	.95 (NA)
Cassareal et al. 1986 (n=972)	IC	Admitting Attitudes	.79 (3)
		Nursing Attitudes	.82 (5)
		Physician Attitudes	.85 (5)
		Housekeeping Attitudes	.85 (5)
LaMonica et al. 1986 (n=100,533)	IC	Technical/Prof.	.81-.85 (14)
		Trust	.84-.90 (18)
		Education	.80-.84 (10)
		Total	.92-.95 (42)

[a] TRT=test-retest reliability estimate; AF=alternate-forms reliability estimate; IC=internal-consistency reliability estimate (see text for definitions).

[b] Authors frequently reported reliability for subscales and global scales that were the sum of two or more subscales. To indicate this is the table, the subscales included in a global scale score are indented and listed immediately before the global scale. (NA) indicates that publication did not specify number of items included in particular scale.

[c] More than one measure of each dimension studied. Number of scales shown in parentheses after scale name; number of items per scale shown, as in other table entries, after reliability estimate.

[d] Scales used by Breslau et al. (1981) identical (except for one item in Technical Quality) to those fielded in Davies et al. (1986; forthcoming); they reported only the range of coefficients across scales (.52 to .82) and noted similarity of estimates for each scale to those reported for the same measures in general populations.

[e] Range across four different insurance plan groups at baseline; similar estimates were reported for each of three subsequent administrations.

[f] Two TRT coefficients reported, first for same interviewer on each administration, second for different interviewers.

al. 1970; Ware et al., 1976). Although many single-item ratings do not meet this minimum reliability standard (Ware et al., 1978), recent work suggests that reliable single-item ratings can be constructed (Ware and Hays, in press).

Only two studies (DiMatteo et al., 1980, 1986) estimated the number of patients required to obtain reliable multi-item scores for individual physicians; 10 patients per physician was inadequate for precise comparisons among individual providers. Too-small numbers of patient per provider is a noteworthy shortcoming of the few quality of care studies published to date that compare individual physicians. Research in progress suggests that 20 to 40 may be required to obtain a reliable quality of care score for each provider.

Validity

The following sections summarize evidence pertinent to the validity of consumers' ratings and reports of features of outpatient and inpatient care as indicators of quality. We organize the evidence according to the feature of care described by the validity variable (i.e., access, interpersonal features, technical process, and, for inpatient care only, overall quality).[2] We found no studies of the validity of consumers' ratings or reports of accessibility or availability of hospital services.

---

2 See Chapter 3, Methods, where we discuss criteria for studies reviewed indepth and discussed in this chapter.

Validity of Consumer Information on Outpatient Care

Access and Availability

Entries in Table 4.2 summarize information about the five studies that met our criteria for indepth review and are relevant to whether consumers' ratings reflect accessibility and availability of outpatient care. Validity variables usually include independent documentation of these features of care.[3] We are not aware of any studies that compared consumers' reports of access-related features (e.g., the usual office hours of providers in the area) with objective evidence regarding those features.[4]

Results indicate that consumers' ratings of overall access (Lurie et al., 1984) and of specific features of access (specifically, resource availability, office and appointment waiting times, waiting time for emergency treatment) are significantly and positively related to independently and objectively observed differences in these particular features of outpatient care.

---

3 A number of studies (e.g., Tessler and Mechanic, 1975; Ware et al., 1976; Aday et al., 1980; Penchansky and Thomas, 1981) that showed close correspondence between consumers' reports of access and availability (e.g., appointment and office waiting times; distance traveled) and their ratings of these features were not included because the reports were not independently obtained.

4 Another approach to assessing access, different from those discussed here but with promise, uses consumers' reports of whether treatment was received for symptoms judged serious or minor by physicians (Shapiro et al., 1986). Short recall periods (e.g., 30 days or less) increase the accuracy of these reports (Aday and Andersen, 1979; Andersen et al., 1979). This methodology has the advantage of taking into account the urgency of care. Supporting their validity, indicators using this approach were sensitive to the effects of differences in generosity of insurance coverage on access to care in a large social experiment (Shapiro et al., 1986).

Table 4.2--Summary of Key Findings, Accessibility of Outpatient Care

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Alpert et al. 1970 (50 pts randomly selected in each of two clinics)	Time-motion study of medical and emergency clinics	Significantly larger percent of pts dissatisfied with waits in clinic with substantially longer waits by time-motion study
Aday et al. 1980 (national sample of noninstitutionalized US population)	Physician and hospital/population ratios by geographic region	Satisfaction with availability of physicians and hospitals significantly lower in areas with fewer resources per capita
Linn 1982 (1418 pts in 20 emergency rooms)	Documentation of time spent waiting for treatment in ER	Pt rating of time spent significantly less favorable with longer time spent
Lurie et al. 1984 (186 medically indigent adults treated at academic medical center)	Documented problems in access for group followed for six mo. after loss in Medicaid benefits and group with no change	Access to care rated less favorably by group that lost benefits and had documented problems in access to care from county
Davies et al. 1986 (1537 non-aged adults sampled from general population)	Objective reports on queuing mechanisms for groups randomized to one HMO or to FFS care from any MD and hospital	Ratings of appointment waits favored FFS and those of office waits favored the HMO, corresponding to known differences in queuing mechanisms

Interpersonal Aspects

Entries in Table 4.3 summarize information from 17 studies that met our review criteria and were relevant to whether data from consumers reflect the interpersonal features of outpatient encounters.[5] Validity variables included experimental manipulation of the provider's behavior toward the patient, and independent observation and classification of the provider's affect during an encounter.

Results indicate that when consumers rate their providers less favorably in terms of interpersonal manner and skills, direct observation will confirm that providers tend to be less familiar with or knowledgeable of the patient, less skilled in dealing with patient feelings, less likely to encourage, support and involve the patient in care, less courteous, less communicative, and more tense and antagonistic in dealing with the patient.

Experimental studies (Corah et al., 1984; Ware et al., forthcoming) indicate that when interpersonal and technical aspects of the provider's behavior are unrelated during an encounter, patients' ratings accurately distinguish different levels of the two, and their ratings of interpersonal features are not influenced by variations in technical process.

---

5 Virtually all studies administered satisfaction measures to collect data from consumers. Although some items in these measures can be considered reports (e.g., "doctors respect their patients' feelings"), most are evaluative statements (e.g., "doctors always do their best to keep the patient from worrying"). Given the predominance of evaluative items in these studies, they shed most light on the validity of consumers' ratings.

Table 4.3.--Summary of Key Findings, Interpersonal Aspects of Outpatient Care

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Bertakis 1977 (100 patients in two studies 1 year apart)	Coding of tape-recordings for information given by physician (explanations, tests, regimen, treatment) & amount retained by patient for experimental & control groups	Ratings of interpersonal and technical quality correlated with amount of information provided by physician actually retained by patient
Stewart et al. 1979 (299 visits to 5 physicians)	Concordance between physician and patient reports of patients' social problems	Satisfaction ratings 3-mo post-encounter unrelated to physician knowledge of social problems
Stiles et al. 1979 (52 patients of 19 physicians in hospital OP clinic)	Coding of physician behavior in 3 segments of interview (history, physical, conclusion) in terms of attentiveness, experience, acknowledgement of other's frame of reference, and focus on others	Ratings of interpersonal behavior positively related to attentiveness in conclusion (r=.43); rating of information-giving related to informativeness in conclusion ratings unrelated to attentiveness during history and physical
DiMatteo et al. 1980 (462 patients, IP and OP, of 71 residents in large community hospital)	Physician scores on objective measures of ability to interpret affective behavior	Ratings of interpersonal aspects of care delivered by residents positively correlated with ability to communicate; ratings of technical quality unrelated to ability
Breslau & Mortimer 1981 (369 parents of disabled children sampled from four clinics)	Continuity of care defined in terms of how frequently parent and child saw same physician	Continuity positively related to ratings of interpersonal care, technical quality, finances, and satisfaction in general; highest correlation with ratings of interpersonal care

Table 4.3.--Summary of Key Findings, Interpersonal Aspects of Outpatient Care (cont'd)

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Smith et al. 1981 (29 new patient interviews with 11 physicians)	Videotapes of interviews scored for length and interaction process	Ratings (content not given) more favorable when physicians gave more information, spent more time discussing prevention; unrelated to amount of agreement, casual conversation, suggestions or opinions
Weinberger et al. 1981 (88 adult OP visits with 20 housestaff)	Videotaped recordings coded for verbal and nonverbal physician behavior	Patient satisfaction higher in encounters with more physician encouragement, coverage of psychosocial issues, and reference to prior visits
Carter et al. 1982 (101 new patient visits)	Trained coders classified dr-pt interaction	Patient satisfaction lower when physicians tense, antagonistic; patient satisfaction linked to physicians' orientation of pt as to what is being done and why
Comstock et al. 1982 (10 adult patients for each of 15 residents)	Trained observers viewed encounters; rated courtesy and information-giving, coded nonverbal behaviors	Post-visit patient satisfaction ratings (chiefly interpersonal items) correlated positively with courtesy, attention, listening, empathy, & information-giving
Bartlett et al. 1984 (60 patients of five residents in primary care residency program)	Trained observers coded videotaped encounters using interpersonal skills scale	Summary satisfaction score (interpersonal skill, information-sharing, quality of care) correlated positively (r=.24) with observer ratings of interpersonal skills

Table 4.3.--Summary of Key Findings, Interpersonal Aspects of Outpatient Care (cont'd)

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Chang et al. 1984 (268 elderly women volunteers assigned randomly to view simulated encounters)	Videotaped patient encounters with nurses and physicians designed to simulate differences in technical quality, patient participation, and handling of psychosocial problems	Overall satisfaction ratings higher for higher levels of all 3 manipulations; ratings of visit length, technical quality, and psychosocial care affected only by technical quality manipulation
Corah et al. 1984 (24 adult patients of two dentists)	Experimental manipulation of dentist-pt interaction in terms of amount said, acceptance, reassurance	Ratings of information/communication and understanding/acceptance more favorable for maximum interaction group; ratings of technical competence were not affected
Stewart 1984 (140 patients of 24 family physicians)	Coders trained in interaction process analysis coded audiotape recordings of physician-pt encounters	Ratings of physician's personal qualities more favorable 10 days post-visit if encounter more patient-centered; ratings of professional competence higher for physicians showing more tension or asking for opinion/help
Carney & Mitchell 1986 (120 1st & 3rd year medical students and 60 simulated patients)	Faculty assessments of student's overall clinical performance used to form two groups (satisfactory & unsatisfactory)	Ratings of student's interpersonal manner by simulated patients were more favorable for students rated satisfactory by faculty
DiMatteo et al. 1986 (239 patients of 28 family practice residents in county hospital; OP and IP)	Physician scores on objective measures of ability to interpret affective behavior	Ratings of affective behavior significantly related (r=.39) to affective ability; ratings of communication and technical quality unrelated to ability

Table 4.3.--Summary of Key Findings, Interpersonal Aspects of Outpatient Care (cont'd)

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Cope et al. 1986 (424 new outpatients of 68 internal medicine residents, large teaching hosp).	Evaluations of physician performance by nurses & by supervising faculty	Patient ratings of physician performance positively correlated with nurse evaluations (r=.33) and with those by supervising faculty (r=.40)
Ware et al. (forth) (109 volunteers randomized to view simulated encounters)	Videotaped encounter with physician designed to simulate differences in interpersonal aspects of care (e.g., warmth, communication style)	Interpersonal aspects rated higher for high than low interpersonal encounters; ratings of technical quality unrelated to interpersonal aspects manipulation

NOTE: OP=outpatient, IP=inpatient.

Technical Process

Ratings.--Entries in the top panel of Table 4.4 summarize information about the six studies relevant to whether consumers' ratings reflect the technical process of outpatient care. Three studies used independent report or manipulation of number and type of services performed as validity variables (Linn, 1975; Sox et al., 1981; Davies et al., 1986); two experimentally manipulated the appropriateness of elements of history-taking and physical examination (Chang et al., 1984; Ware et al., forthcoming); one (Linn, 1982) used independent judgements of technical quality.

These studies offer preliminary answers to three questions that have been raised regarding the validity of consumers' ratings of technical process: Can consumers distinguish care judged technically good vs. poor by physicians? Are consumers "seduced" by the kind or number of procedures into believing that services provided were appropriate? Does the provider's interpersonal manner interfere with the consumer's accurate assessment of technical process?

Consumers' ratings of technical quality appear to be somewhat inflated relative to those made by physicians (Ware et al., forthcoming). Despite this, evidence from two experiments in which manipulations of technical process were verified and rated by physicians suggests that, at least for common problems (e.g., chest pain in elderly patients; upper respiratory infection), consumers' ratings of the completeness and thoroughness of care accurately distinguish between encounters for which technical process was judged good and less-than-good by physicians (Chang et al., 1984; Ware et al., forthcoming). Findings from the Linn (1982) study indicate that consumers' ratings of the overall quality of care are also sensitive to documented variations in technical process.

Table 4.4--Summary of Key Findings, Technical Process of Outpatient Care

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Ratings		
Linn 1975 (1739 encounters in 11 outpatient facilities	Number and type of services performed (e.g. history, exam, lab tests, X-rays)	Summary satisfaction score unrelated to number or type of services performed
Sox et al. 1981 (176 outpatients in VA hospital with chest pain	Performance or nonperformance of diagnostic tests	Those receiving tests rated care for chest pain better than usual, were less likely to feel that too few tests were done; no differences in ratings of interpersonal care and communication
Linn 1982 (1418 patients in 20 emergency rooms)	Technical proces of burn care judged against clinical algorithm	Pt ratings of overall ER care significantly less favorable with more deviations from algorithm
Chang et al. 1984 (268 elderly women volunteers assigned randomly to view simulated encounters)	Videotaped patient encounters with nurses and physicians designed to simulate differences in technical quality (whether relevant medical history and physical examination items were performed)	Overall satisfaction rated higher for high than low technical encounters; satisfaction with visit length, technical quality, and psychosocial care higher for high than low technical encounters
Davies et al. 1986 (1537 non-aged adults sampled from general population)	3- to 5-yr followup of groups randomized to HMO or FFS care; expenditures on use 25% lower at HMO	All but low-income, initially well subgroup rated technical quality of FFS care more favorably than HMO care
Ware et al. (forth) (109 volunteers randomized to view simulated encounters)	Videotaped encounter with physician designed to simulate differences in technical quality (whether relevant medical history and physical examination items were performed)	Technical quality rated higher for high than low technical encounters; ratings of interpersonal features unrelated to technical quality manipulation

Table 4.4--Summary of Key Findings, Technical Process of Outpatient Care (cont'd)

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Reports		
Gerbert and Hargreaves 1987 (214 COPD patients of 63 physicians)	Physician reports of technical elements of the outpatient visit	Agreement between physician and patient reports on: tests ordered, 96%; treatments mentioned, 94%; occurrence of patient education, 88%
Gerbert et al. (in press) (197 COPD patients of 83 physicians)	Videotaped outpt visits checked for mention of theophylline prescription	Pt report on interview of having prescription in strong agreement with videotaped observation (kappa=0.50, $p < .001$)
Ware et al. (forth) (109 volunteers randomized to view simulated encounters)	Videotaped encounter with physician designed to simulate differences in technical quality (whether relevant medical history and physical examination items were performed)	Technical quality rated higher for high than low technical encounters; ratings of interpersonal features unrelated to technical quality manipulation

Table 4.5--Summary of Key Findings, Interpersonal Aspects of Inpatient Care

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Ley et al. 1976 (63 inpatients at hospital in Great Britain)	Random groups experiment of extra MD visit to assess, aid pt understanding	Experimental group pts rated communication significantly higher than controls (no visit) or placebo group (visit, no information content)
Hinshaw et al. 1983 (88 surgical pts)	Random groups experiment of perioperative RN visits to reassure and educate; quality independently judged better for visited patients	Pt ratings of trusting relationship and technical/professional nursing care significantly more favorable for visited group
Kane et al. 1985 (246 inpatients in VA hospital)	Random groups expt of hospice ward/team intervention that increased provider communication, more pt/family involvement in care [a]	Pt ratings of involvement in care and care overall (technical, interpersonal, general) significantly higher for hospice group; ratings of physical environment unaffected by intervention

[a] Manipulation check not reported by authors.

Findings from two (Sox et al., 1981; Davies et al., 1986) studies suggest that consumers' ratings of technical process do reflect, at least in part, how many services they received; those reported by Linn (1975) showed no relationship. Results from two experiments (Corah et al., 1984; Ware et al., forthcoming) and observational studies (DiMatteo et al., 1980, 1986) indicated that the physician's interpersonal manner had an insignificant effect on consumers' ratings of technical process.[6]

Reports.--Entries in the bottom panel of Table 4.4 summarize information about the three studies relevant to the issue of whether consumers' reports accurately reflect elements of the technical process of outpatient care. Results illustrate the relatively high accuracy of consumers' reports regarding elements of outpatient care. Consumers in the Ware et al. (forthcoming) experiments identified medical history and physical examination items that were and were not done with 70 to 88 percent accuracy; better-educated respondents were more accurate. COPD patients were very accurate (when compared with physicians) in reporting tests ordered (96 percent), treatments mentioned (94 percent), and occurrence of patient education (88 percent) (Gerbert and Hargreaves, 1987), and in reporting prescription medications (when compared with data from videotaped encounters) (Gerbert et al., in press).

Validity of Consumer Information on Inpatient Care

Interpersonal Aspects

Entries in Table 4.5 summarize information from the three studies that met our review criteria and were relevant to whether data from consumers

---

6 Results from these four studies were summarized in Table 4.3, because the validity variables related to interpersonal aspects of care.

reflect the interpersonal features of inpatient care. All three were experiments; interventions focused on modifying aspects of provider behavior toward patients by medical staff (Ley et al., 1962), nursing staff (Hinshaw et al., 1983), and both (Kane et al., 1985).

All studies provided convergent evidence of validity. Inpatients' ratings of the interpersonal features of care that were manipulated experimentally (e.g., communication, involvement in care) were significantly higher for the groups that received the interventions. Ratings of technical/professional aspects of nursing care (Hinshaw et al., 1983) and of care overall (Kane et al., 1985) were also sensitive to these interventions.

Technical Quality

The single entry in Table 4.6 summarizes information related to whether consumer data reflects technical features of inpatient care. Because this is the only study, we relaxed our criteria somewhat to include it; given known gaps in medical records as a source of quality-related data the validity variable they used (record-based judgments of technical process against implicit criteria), is not the best standard against which to test patient ratings. Findings from the Ehrlich et al. (1961) study indicate that consumers' overall judgments of quality of medical care delivered during hospital episodes were inflated relative to those made by physicians, but were more likely to be favorable if care was judged good (as opposed to less-than-good) by the physicians.

Overall Quality

Ratings.--Entries in the top panel of Table 4.7 summarize information from the two studies that met our review criteria and were relevant to whether

Table 4.6--Summary of Key Findings, Technical Process of Inpatient Care

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Ehrlich et al. 1961 (283 Teamsters in 105 NY hospitals)	MD judgments of technical quality based on record review	Pt ratings of medical care while hospitalized significantly related to MD judgments of technical quality: 5 of 6 stays rated not good by pts judged fair or poor by MDs; significantly more patients judged care best when rated excellent or good by MDs (86% vs. 74%)

Table 4.7--Summary of Key Findings, Overall Quality of Inpatient Care

Author/Year (Sample)	Validity Variable(s)	Summary of Findings
Ratings		
Rice et al. 1963 (457 psychiatric inpatients)	Sum of staff rankings of wards on 7 criteria [a]	Pt ratings of overall hospital care (sum of physical facilities, pt services, and pt management) identical staff rankings
Hinshaw et al. 1982 (88 surgical pts)	Number of care recommendations made by RN (more presumed to indicate better quality RN care)	Pt ratings of trust, technical, education, and overall hospital care significantly more favorable when more recommendations made
Reports		
Abdellah & Levine 1957 (60 inpatients at a Midwestern hospital)	Query of staff member to determine whether reported omission in nursing care had occurred	100% of pt-reported omissions in nursing care verified by staff report
Abdellah & Levine 1958 (>9000 inpatients in 60 Midwestern hospitals)	Staffing levels of professional nurses (higher levels presumed to indicate better quality of care)	Significantly fewer pt-reported omissions in nursing therapy with more professional RNs; staffing levels weakly or not related to reported omissions in environmental features or attention to dietary needs
Ehrlich et al. 1961 (283 Teamsters in 105 NY hospitals)	Chart review to identify diagnostic tests prior to a hospitalization	Tendency to under-report: 1/3 of patients failed to report tests mentioned in chart

[a] Criteria included: adequacy of physical facility; crowdedness; patient morale; staff morale; amount of staff/patient contact; degree of patient/staff harmony; and amount of freedom granted patients.

consumers' ratings reflect the overall quality of inpatient care. Validity variables included summary rankings of psychiatric wards by staff on a range of criteria (Rice et al., 1963) and recommendations for care made by nurses (Hinshaw et al., 1982).

Results from both studies support the validity of consumers' ratings of overall quality. Rice et al. (1963) noted that patients' overall ratings ranked psychiatric wards identically to the rankings made by staff. In the Hinshaw et al. (1982) study, patients' ratings of overall quality were significantly higher when nurses made more recommendations regarding care; they presumed that more recommendations reflected better quality nursing care.

Reports.--Entries in the bottom panel of Table 4.7 summarize information from the three studies that met our review criteria and were at all relevant to whether consumers' reports reflect the overall quality of inpatient care. Validity variables included staff reports of omissions in nursing care (Abdellah and Levine, 1957), staffing levels of professional nurses (Abdellah and Levine, 1958), and chart reviews (Ehrlich et al., 1961). Results provide an equivocal answer to the question of whether consumers' reports are sensitive to overall quality, in part because none of the studies had well-defined validity variables and in part because two (Abdellah and Levine, 1957; Ehrlich et al., 1961) of the three reported results in such a way that true rates of under- (or over-) reporting could not be discerned.

Abdellah and Levine (1957) reported that 100 percent of omissions in nursing care (e.g., failure to administer medications on schedule; failure to answer call bell) reported by inpatients were verified by staff. Because staff were asked to verify only those omissions reported by patients, we cannot be sure that under-reporting did not occur.

Their later study (Abdellah and Levine, 1958) illustrates the

sensitivity of inpatients' reports of aspects of nursing care for which professional nurses (RNs) were expected to be more responsible (e.g., therapy) to higher levels of professional nursing staff. By contrast reports about things for which RNs were less responsible (e.g., attention to dietary needs) were unrelated to professional/nonprofessional nurse staffing ratios.

Ehrlich et al. (1961) found that a substantial minority (one-third) of patients under-reported the diagnostic tests they had prior to a hospitalization. Given the way they reported their data, we cannot determine from this study the number or type of tests that were under-reported.

Feasibility

There are three basic questions regarding the feasibility of obtaining data from consumers on quality-related attributes of care: (1) whether appropriate questionnaires and data collection techniques are available and/or can be developed; (2) whether potential respondents can be identified and will agree to respond; and (3) whether the costs of obtaining data from consumers are reasonable.

Each can be answered affirmatively. In part because the literature we have reviewed here directly addresses only the second question (i.e., by reporting response rates), these answers depend heavily on our own practical experiences and knowledge of the considerable literature on survey research methods in general, a detailed synthesis of which was beyond the scope of this review. We briefly summarize the evidence behind our affirmative responses below.

Availability of Data Collection Instruments and Techniques

Our review indicates that there are several good batteries of items

that can be used to obtain consumers' ratings of outpatient care, where "good" means that the batteries do a comprehensive job of representing one or more attributes of care for which consumers provide valid data (e.g., see scales and/or batteries cites in Tables 4.2-4.4). Published batteries for obtaining ratings of inpatient care rarely have such characteristics, because we have considerable less information about the dimensions of hospital care that relate to quality (Rubin et al., forthcoming), and because fewer studies have examined the validity of published batteries. An ongoing collaborative effort by the Hospital Corporation of America, Harvard Community Health Plan, and us to develop and test an hospital satisfaction survey should prove useful information in this regard (Berwick et al., forthcoming).

A wide variety of data collection techniques are available and have been used to obtain information from consumers. Self- and interviewer-administration (usually in-person; sometimes by telephone) are the most commonly used. The "best" technique will vary depending on the study population, complexity of the data collection instrument, and a variety of other considerations. We recommend involvement of reputable survey research unit or firm early in the design and planning of data collection.

Availability of Willing Respondents

The management information systems available in many outpatient and inpatient settings, provide a ready way of identifying potential respondents to consumer surveys. Depending on the focus of a particular quality of care evaluation, their data might be manipulated to identify, for example, a universe of consumers (in the case of an enrolled sample); users vs. non-users; those who complain or lodge formal grievances; hospitalized patients, by admitting diagnosis, procedure, and/or unit; and patients who see a

particular provider.

Because consumers are generally willing to discuss their medical care experiences and attitudes, good response rates (i.e., 70 percent or higher) can be achieved on consumer surveys (for example, Ware et al., 1976; French, 1981; Abramowitz et al., 1987; Brook et al., 1987; Berwick et al., forthcoming). Lower response rates, which raise questions about sample bias, are often caused by inadequate follow-up efforts.

Costs of Obtaining Consumer Data

Few published studies include any information about the costs of collecting data on quality of care from patients or other, more traditional, sources (e.g., medical record audit, computerized claims audit). Survey costs will vary markedly depending on such factors as administration method, dispersion of the sample, availability of potential respondents, follow-up procedures, and questionnaire length; mail and telephone surveys usually cost markedly less than personal interviews (Fowler, 1984).

What little evidence is available suggests that acquiring information from consumers costs no more, and in most circumstances less, than that obtained from records. Recently obtained estimates for medical record abstractions designed for quality of care evaluation ranged from $35 to $55 per record (Davies and Ware, in press); costs for mail and telephone surveys of typical length (e.g., 15-20 minutes) appear to range from about $12 to $15 dollars (Walker and Restuccia, 1984; Davies and Ware, in press; Berwick et al., forthcoming).

Whether information from consumers provides a cost-effective method of obtaining data on quality of care indicators is an open question. We identified only one study (Gerbert and Hargreaves, 1987) that compared the

costs of obtaining quality-relevant data from different sources (i.e., physician and patient interviews, medical record abstraction, coding of videotaped encounters). Findings from that study (Gerbert and Hargreaves, 1987; Gerbert et al., in press) illustrate forcefully the need to regard data on technical process of care from all sources as complements, rather than as substitutes, until we can better identify which source provides the most accurate (and least expensive) information regarding different aspects of technical process. Given the paucity of data from other traditional sources regarding the interpersonal aspects of care, and the intrusiveness, complexity, and cost of using such approaches as direct observation and coding of the provider-patient encounter (Inui and Carter, 1985), obtaining information from consumers of care appears to be the most cost-effective approach.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

We set out to examine the suitability of consumers' assessments of physician and hospital performance as indicators of the quality of care. At the outset, we made several distinctions that proved to be important, including distinctions between: (1) assessments of care in outpatient and inpatient settings; (2) specific quality of care attributes (access to care, interpersonal manner, and technical process); and (3) consumers' ratings versus reports. Our findings and conclusions regarding the suitability of consumer assessments occasionally differ along the lines of these distinctions. On the basis of our review, we conclude that it is possible to construct valid consumer-based indicators of quality of care and that there are good reasons to use them given shortcomings inherent in alternative strategies.

There are good reasons to rely on consumer assessments of access to care in assessing the quality of physician and hospital performance. Although the volume of evidence on this issue is less than that for interpersonal quality, it is quite strong. We found no published evidence that warrants questioning of the validity of consumers' assessments of access to care in the outpatient setting. Available evidence comes from comparisons with results from time-and-motion studies, objective assessments of the availability of medical care resources, and interventions designed to alter access. Results from the five studies that met our selection criteria clearly indicate that when patients evaluate access more or less favorably, corresponding differences in access to care in fact are very likely to exist. Numerous

examples of indirect evidence (i.e., studies linking consumer reports to consumer evaluations of care) were also found. We found no studies of the validity of consumers' assessments of access to hospital services.

By all standards considered, there is a strong case for using consumer assessments as indicators of the interpersonal aspects of care. The case is strong for evaluations of both the interpersonal manner of physicians in outpatient settings and of the manner of physicians and hospital staff in inpatient settings. This conclusion is based on several considerations. First, we found no practical or valid alternative source of information on interpersonal manner described in the literature. We eliminated direct observation on grounds of impracticality (intrusiveness, complexity of coding schemes, and expense), and concerns about whether ratings by trained observers adequately reflect the patient's values. We found no evidence that the medical record, in either outpatient or inpatient settings, provides valid information about the interpersonal aspects of care. Even if providers routinely made notes about the quality of their interpersonal relationships with patients, there would still be reason to question the validity of such notes. Who is more qualified than the individual consumer to judge the interpersonal manner of physicians and other health care providers in light of consumers' standards?

These arguments themselves, however, provide no guarantee that consumer assessments are valid indicators of the interpersonal quality of care. The crucial pieces of the puzzle, pulled together for the first time in this report, are published findings regarding the empirical validity of consumers' ratings of the interpersonal aspects of care. We found considerable evidence that these ratings are valid indicators of interpersonal aspects of care in

outpatient and inpatient settings. The evidence across settings comes from 20 studies that compared results for objective measures of the interpersonal aspects of care with those for patients' ratings. These validation standards included direct observation by trained observers, evaluations by physicians and other health care providers, analyses of audio- and videotape recordings, randomized-groups experiments to evaluate interventions designed to change the interpersonal aspects of care, and studies of randomized groups in which variations in interpersonal aspects of care were experimentally manipulated. Of the 26 outpatient studies that satisfied selection criteria, 17 yielded evidence in support of the validity of consumer assessments of interpersonal quality. Of the 8 studies fielded in inpatient settings, 3 yielded evidence in support of the validity of consumer assessments of interpersonal quality.

We found relatively little published evidence regarding the validity of consumers' assessments of technical process. This is unfortunate, because most common methods for assessing technical quality carry high dollar and time costs. For example, medical record abstraction according to protocol requires professional input to the development of algorithms, highly-trained personnel (e.g., nurses, medical record librarians, physicians) to do the abstraction, and depends on the ready availability of the record itself. Further, the outpatient record (as opposed to the inpatient record) is an incomplete source of information about the quality of the technical process. The search for data sources that may cost less and that help to fill the gaps leads some to consider surveying consumers about their care.

The available evidence, although limited and only from outpatient settings, generally supports the use of consumer ratings of the technical process as indicators of the quality of care. Specifically, those few studies that have verified differences in technical process (e.g.,

physician/staff assessments, experimental manipulations) and have compared results with patient ratings have consistently linked the two. Moreover, evidence from experimental studies suggests that, at least for relatively common outpatient conditions, the physician's interpersonal manner does not obscure the consumers' ability to detect variations in technical process. Thus, the available evidence does not support the strong editorial comments that have been made against using consumers' ratings for this purpose. However, pending further research on this issue and replication of these findings, we recommend that consumers ratings of technical process be used only in conjunction with highly credible data about that technical process for purposes of evaluating the quality of care.

Our findings suggest that a promising but rarely employed strategy for consumer-based quality of care assessments would be one based on consumers' reports of what does and does not occur. This approach makes no assumption about their qualifications as judges, only about the accuracy of their reports. Physicians and/or those using algorithms for evaluating technical process can use these reports to make the actual judgment regarding quality of the technical process. We recommend further research to determine those aspects of the technical process that can and cannot be reported accurately by consumers in order to test the suitability of this strategy in both outpatient and inpatient settings.

Not surprisingly, available evidence establishes a direct link between the specificity of the content of consumer quality assessments and their validity. Technical process, interpersonal process, and access to care are distinct quality of care attributes that can be measured and interpreted separately. Validity is generally better when there is a good match between

the content of the assessment and the quality attribute of interest. However, more global measures (e.g., overall satisfaction ratings, whether patients are willing to recommend a hospital to others, health care plan disenrollment rates) are not unrelated to quality of care. Given the overriding importance of quality of care to consumers, large differences in such global indicators of satisfaction are likely to reflect differences in quality. However, because of their sensitivity to a wide range of differences, other interpretations of such indicators should be kept in mind. Further, global measures are not as programmatically useful because they do not provide clues as to which attributes of quality are most likely to account for any differences that are observed.

Priorities for future research in the inpatient setting should include studies of consumer assessments of specific features of quality of care that have not been distinguished in work to date, including interpersonal and technical aspects of medical and nursing care, information giving and other aspects of communication, and patient and family involvement in care. Little little is known about how differences in the quality of the inpatient technical process are experienced by patients or how they affect patient assessments. Our review has corrected one common misconception, namely, that consumer assessments of the quality of hospital care are determined by the amenities. To the contrary, patient judgements of the overall quality of hospital care seem to be determined most by the quality of nursing and medical care. Finally, studies of how variations in process and patient experiences are related to subsequent patient outcomes are needed to address quality of care most directly, namely, the outcome question.

In evaluating the validity of consumers' assessments, we examined the content of published indicators to determine how well they reflected consumers' values, and their comprehensiveness in relation to the universe of consumer experiences. Although a number of published instruments are quite comprehensive, none covers all attributes of quality well. Available taxonomies of consumer experiences with outpatient (Ware et al., 1978) and inpatient care (Rubin et al., forthcoming) should be used as minimum standards for judging the content of candidate measures. Published batteries designed to obtain data from consumers about inhospital care are particularly lacking in this regard, and further developmental work is required.

It is likely that quality considerations will be increasingly emphasized in attempts to market prepaid and other group plans, health insurance benefits, and hospital facilities to consumers. Such efforts appeal directly to consumers' desires for "quality" health care. This trend underscores both the potential value of published consumer-based information regarding the quality of physician and hospital performance and the potential for abuse of the data. Clearly, valid indicators should be favored over invalid indicators. Because of the importance of measurement method and consumer sampling in determining results, there is a need for standardization of measurement and sampling methods and development of minimum standards for reporting results to the public.

For these reasons, we recommend that careful attention be given to methodological issues that are likely to influence conclusions based on consumer assessments of quality. For comparisons between physicians or hospitals to be valid, they must be based on standardized instruments, data collection procedures (e.g., personal or telephone interview, self-

administered questionnaire), survey methods (e.g., timing of administration), and representative samples. Reproducible scores can be achieved only if methods are carefully standardized.

We conclude with a final argument for routinely and carefully monitoring consumer-based indicators of the quality of physician and hospital care, which makes no assumptions about the validity of these indicators in relation to the realities of the medical care process. Instead, the argument is based on strong empirical evidence that consumers' perceptions of quality of care influence their behavior. The behaviors affected include doctor-shopping, complaints, disenrollment, compliance, and utilization of services. Such behaviors have noteworthy consequences to their health and the quality of their care.

REFERENCES

Abdellah, F.G., Levine, E., "Developing a measure of patient and personnel satisfaction with nursing care," Nursing Research 5:100-108, 1957.

Abdellah, F.G., Levine, E., Effect of Nurse Staffing on Satisfaction with Hospital Care. Hospital Monograph Series No. 4. (Chicago: American Hospital Association, 1958).

Abramowitz, S., Cote, A.A., Berry, E., "Analyzing patient satisfaction: A multianalytic approach", Quality Review Bulletin 13:122-130, 1987.

Aday, L.A., Andersen, R.A., Access to Medical Care. (Ann Arbor: Health Administration Press, 1975).

Aday, L.A., Andersen, R.A., Fleming, G.V., Health Care in the U.S.: Equitable for Whom?, (Beverly Hills, CA: Sage Publications, 1980).

Alpert, J.J., Kosa, J., Haggerty, R.J., et al., "Attitudes and satisfactions of low-income families receiving comprehensive pediatric care," American Journal of Public Health 60:499-506, 1970.

American Board of Internal Medicine, "Evaluation of humanistic qualities in the internist," Annals of Internal Medicine 99:720-724, 1983.

Andersen, R., Kasper, J., Frankel, M.R., et al., Toral Survey Error, (San Francisco: Jossey-Bass Publishers, 1979).

Bartlett, E.E., Grayson, M., Barker, R., et al., "The effects of physician communications skills on patient satisfaction, recall, and adherence," Journal of Chronic Diseases 37:755-764, 1984.

Becker, M.H., Drachman, R.H., Kirscht, J.P. "A new approach to explaining sick-role behavior in low-income populations," Journal of Public Health 64:205-216, 1974.

Bertakis, K.D. "The communication of information from physician to patient: A method for increasing patient retention and satisfaction," The Journal of Family Practice 5:217-222, 1977.

Berwick, D., et al., "Development of a patient hospital satisfaction questionnaire," (forthcoming).

Breslau, N., Mortimer, E.A., "Seeing the same doctor: Determinants of satisfaction with specialty care," Medical Care 19:741-758, 1981.

Brook, R.H., Avery, A.D., "Quality assurance in the United States: From there to where?," A Question of Quality: Roads to Assurance in Medical Care, G. McLachlan, (ed.) (London: Oxford University Press, 1976).

Brook, R.H., Avery, A.D., "Quality assessment: Issues of definition and measurement," Medical Peer Review Systems: Theory and Practice, P.Y. Ertel and M.G. Aldridge (eds.) (St. Louis, Mosby and Company, 1977).

Brook, R.H., Fink, A., Kosecoff, J., et al., "Educating physicians and treating patients in the ambulatory setting: Where are we going and how will we know when we arrive?", Annals of Internal Medicine 107:392-398, 1987.

Brook, R.H., Ware, J.E., Rogers, W.H., et al., "Does free care improve adults' health?: Results from a randomized controlled trial," New England Journal of Medicine 309:1426-1434, 1983.

Carmel, S. "Satisfaction with hospitalization: A comparative analysis of three types of services," Social Science and Medicine 21:1243-1249, 1985.

Casarreal, K.M., Mills, J.I., Plant, M.A., "Improving service through patient surveys in a multihospital organization," Hospital and Health Services Administration 31:41-52, 1986.

Carney, S.L., Mitchell, K.R., "Satisfaction of patients with medical students' clinical skills," Journal of Medical Education 61:374-379, 1986.

Carter, W.B., Inui, T., Kukull, W., et al., "Outcome-based doctor-patient interaction analysis: II. Identifying effective provider and patient behaviors," Medical Care 20:550-566, 1982.

Chang, B.L., Uman, G.C., Linn, L.S., et al., "The effect of systematically varying components of nursing care on satisfaction in elderly ambulatory women," Western Journal of Nursing Research 6:367-379, 1984.

Christoffel, T., Loewenthal, M., "Evaluating the quality of ambulatory care: A review of emerging methods," Medical Care 15:877-897, 1977.

Comstock, L.M., Hooper, E.M., Goodwin, J.M., et al., "Physician behaviors that correlate with patient satisfaction," Journal of Medical Education 57:105-112, 1982.

Cope, D.W., Linn, L.S., Leake, B.D., et al., "Modification of residents' behavior by preceptor feedback of patient satisfaction," Journal of General Internal Medicine 1:394-398, 1986.

Corah, N.L., O'Shea, R.M., Pace, L.F., et al., "Development of a patient measure of satisfaction with the dentist: The Dental Visit Satisfaction scale," Journal of Behavioral Medicine 7:367-373, 1984.

Davies, A.R., Ware, J.E., "Involving consumers in quality of care assessment: Do they provide valid information?," Health Affairs, in press.

Davies, A.R., Ware, J.E., Brook, R.H., et al., "Consumer acceptance of prepaid and fee-for-service care: Results from a randomized controlled trial," Health Services Research 21:429-452, 1986.

Davies, A.R., Ware, J.E., Brook, R.H., et al., Consumer Acceptance of Prepaid and Fee-for-service Care: Results from a Randomized Controlled Trial," R-3219-HHS. (Santa Monica, CA: The RAND Corporation, forthcoming).

DiMatteo, M.R., Hays, RD., Prince, L.M. "Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload," Health Psychology 5:581-594, 1986.

DiMatteo, M.R., Taranta, A., Friedman, H.S., et al., "Predicting patient satisfaction from physicians' nonverbal communication skills," Medical Care 18:376-387, 1980.

Doering, E.R., "Factors influencing inpatient satisfaction with care," Quality Review Bulletin 9:291-9, 1983.

Donabedian, A., "Evaluating the quality of medical care," Milbank Memorial Fund Quarterly 44(part 2):166-203, 1966.

Donabedian, A., Explorations in Quality Assessment and Monitoring. Volume I: The Definition of Quality and Approaches to its Measurement, (Ann Arbor, MI: Health Administration Press, 1980).

Ehrlich, J., Morehead, M.A., Trussell, R.E., The Quantity, Quality, and Costs of Medical and Hospital Care Secured by a Sample of Teamster Families in the New York Area, (New York: Columbia University School of Public Health and Administrative Medicine, 1961).

Franklin, F.J., McLemore, S.D., "A scale for measuring attitudes toward student health services," Journal of Psychology 66:143-147, 1967.

Franklin, F.J., McLemore, S.D., "Factors affecting the choice of medical care among university students," Journal of Health and Social Behavior 11:311-319, 1970.

French, K., "Methodologic considerations in hospital patient opinion surveys," International Journal of Nursing Studies 18:7-32, 1981.

Fowler, F.J., Survey Research Methods, Applied Social Research Methods Series, Vol. 1, (Beverly Hills, CA: Sage Publications, 1984).

Gerbert, B., Hargreaves, W.A., "Measuring physician behavior," Medical Care 24:838-847, 1987.

Gerbert, B., Stone, G., Stulbarg, M., et al., "Agreement among physician assessment methods: Searcing for the truth among fallible methods," Medical Care (in press).

Greenley, J.R., Schulz, R., Nam, S.H., et al., "Patient satisfaction with psychiatric inpatient care: issues in measurement and application," Research in Community Mental Health 4:303-319.

Helmstadter, G.C., Principles of Psychological Measurement, (New York, Appleton-Century-Crofts, 1964).

Hinshaw, A.S., Atwood, J.R., "A patient satisfaction instrument: Precision by replication," Nursing Research 31:170-191, 1982.

Hinshaw, A.S., Gerber, R.M., Atwood, J.R., et al., "The use of predictive modeling to test nursing practice outcomes," Nursing Research 32:35-42, 1983.

Hulka, F.S., Zyzanski, S.H., Cassel, J.C., et al., "Scale for the measurement of attitudes toward physicians and primary medical care," Medical Care 8:429-436, 1970.

Inui, T.S., Carter, W.B., "Problems and prospects for health services research on provider-patient communication," Medical Care 23:521-538, 1985.

Kane, R.L., Klein, S.J., Bernstein, L., et al., "Hospice role in alleviating the emotional stress of terminal patients and their families," Medical Care 23:189-197, 1985.

Kisch, A., Reeder, L., "Client evaluation and physician performance," Journal of Health and Social Behavior 10:51-58, 1969.

LaMonica, E.L., Oberst, M. T., Madea, A.R., et al., "Development of a patient satifaction scale," Research in Nursing and Health 9: 43-50, 1986.

Lebow, J., "Consumer assessments of the quality of medical care," Medical Care 12:328-337, 1974.

Ley, P., Kinsey, J., Atherton, S.T., "Increasing patients' satisfaction with communication," British Journal of Social and Clinical Psychology 15:403-413, 1976.

Linn, L.S., Greenfield, S., "Patient suffering and patient satisfaction among the chronically ill," Medical Care 20:425-431.

Linn, L.S., "Factors associated with patient evaluation of health care," Milbank Memorial Fund Quarterly 53:531-548, 1975.

Lurie, N., Ward, N.B., Shapiro, M.F., et al., "Termination from Medi-Cal: Does it affect health?," New England Journal of Medicine 311:480-484, 1984.

Lurie, N., Ward, N.B., Shapiro, M.F., et al., "Termination of Medi-Cal benefits: A follow-up study one year later," New England Journal of Medicine 314:1266-1268, 1986.

Lyons, R.F., Payne, B.C., "The relationship of physicians' medical recording performance to their medical care performance," Medical Care 12:463-469, 1974.

Marquis, M.S., Davies, A.R., Ware, J.E., "Patient satisfaction and change in medical care provider: A longitudinal study," Medical Care 21:821-829, 1983.

Murray, J.P., "A comparison of patient satisfaction among prepaid and fee-for-service patients," The Journal of Family Practice 24:203-207, 1987.

Nunnally, J.C., Psychometric Theory, Second edition. (New York: McGraw Hill Book Company, 1978).

Penchansky, R., Thomas, J.W., "The concept of access: Definition and relationship to consumer satisfaction," Medical Care 19:127-140, 1981.

Rice, C.E., Berger, D.G., Klett, S.L., et al., "The ward evaluation scale," Journal of Clininical Psychology 19:251-258, 1963.

Rojek, D.G., Clemente, F., Summers, C.F., "Community satisafaction: A study of contentment with local services," Rural Sociology 40:177-192, 1975.

Rosenberg, G., Speedling, E.J., Rehr, H., et al., "Some effects of a hospital employee strike on patient satisfaction," Mount Sinai Journal of Medicine (NY) 52:259-264, 1985.

Roter, D.L., "Patient participation in the patient-provider interaction: The effects of patient question asking on the quality of interaction, satisfaction and compliance," Health Education Monographs 5:281-315, 1977.

Rubin, H.R., et al., "Patient evaluations of inpatient care: A review of the literature," (forthcoming).

Shapiro, M.F., Ware, J.E., Sherbourne, C.D., "Effects of cost sharing on seeking care for serious and minor symptoms: Results of a randomized controlled trial," Annals of Internal Medicine 104:246-251, 1986.

Smith, C.K., Polis, E., Hadac, R.R., "Characteristics of the initial medical interview associated with patient satisfaction and understanding," The Journal of Family Practice 12:283-288, 1981.

Souelem, O., "Mental patients' attitudes toward mental hospitals," Journal of Clinical Psychology 7:329-337, 1979.

Sox, H.C., Jr., Margulies, I., Sox, C.H. "Psychologically mediated effects of diagnostic tests," Annals of Internal Medicine 95:680-685, 1981.

Standards for Educational and Psychological Tests, (Washington, DC: American Psychological Association, 1974).

Stewart, M.A., "What is a successful doctor-patient interview? A study of interactions," Social Science and Medicine 19:167-175, 1984.

Stewart, M.A., McWhinney, I.R., Buck, C.W., "The doctor/patient relationship and its effect upon outcome," Journal of the Royal College of General Practitioners 29:77-82, 1979.

Stiles, W.B., Putnam, S.M., James, S.A., et al., "Dimensions of patient and physician roles in medical screening interviews," Social Science and Medicine 13A:335-341, 1979.

Stirewalt, C.F., Linn, M.S., Godoy, G., et al., "Effectiveness of an ambulatory care telephone service in reducing drop-in visits and improving satisfaction with care," Medical Care 20:739-748, 1982.

Tessler, R., Mechanic, D., "Consumer satisfaction with prepaid group practice: A comparative study," Journal of Health and Social Behavior 16:95-113, 1975.

Thompson, H.C., Osborne, C.E., "Office records in the evaluation of quality of care," Medical Care 14:294-314, 1976.

Walker, A.H., Restuccia, J.D., "Obtaining information on patient satisfaction with hospital care: Mail vs. telephone," Health Services Research 19:291-306, 1984.

Ware, J.E., Brook, R.H., Rogers, W.H., et al., "How do health outcomes at an HMO compare with those of fee-for-service care? Results from a randomized controlled trial," Lancet 1:1017-1022, 1986.

Ware, J.E., Curbow, B., Davies, A.R., et al., Medicaid Satisfaction Surveys, 1977-1980: A Report of the Prepaid Health Research, Evaluation, and Demonstration Project, (Sacramento, CA: California State Department of Health Services, 1981).

Ware, J.E., Davies, A.R., "Behavioral consequences of consumer dissatisfaction with medical care," Evaluation and Program Planning 6:291-297, 1983.

Ware, J.E., Davies-Avery, A., Stewart, A.L., "The measurement and meaning of patient satisfaction," Health & Medical Care Services Review 1:1-15, 1978.

Ware, J.E., Hays, R., "Methods for measuring patient satisfaction with specific medical encounters," Medical Care (in press).

Ware, J.E., Kane, R.L., Davies, A.R., et al., "The patient's role in assessing the quality of the medical care process," (forthcoming).

Ware, J.E., Snyder, M.K., "Dimensions of patient attitudes regarding doctors and medical care services," Medical Care 13:669-682, 1975.

Ware, J.E., Snyder, M.K., Wright, W.R., Development and Validation of Scles to Measure Patient Satisfaction with Health Care Services: Vol. I of a Final Report. Part B: Results Regarding Scales Constructed from the Patient Satisfaction Questionnaire and Measures of Other Health Care Perceptions, NTIS Publ. No. PB 288-330, (Springfield, VA: National Technical Information Service, 1976).

Ware, J.E., Wright, W.R., Snyder, M.K., et al., "Consumer perceptions of health care services: Implications for academic medicine," Journal of Medical Education 50:839-848, 1975.

Weinberger, M., Greene, J.Y., Mamlin, J.J., "The impact of clinical encounter events on patient and physician satisfaction," Social Science and Medicine 15E:239-244, 1981.

Zyzanski, S.J., Hulka, B.S., Cassel, J.C., "Scale for the measurement of "satisfaction" with medical care," Medical Care 12:611-620, 1974.